Multiclass Food Classification

Hussein Younes
h.younes@innopolis.university

Dinar Zayahov
d.zayahov@innopolis.university

Abstract

In this project, we shall create a system for food multi-class classification. The system uses a convolution-based architecture to classify food images. We are also going to deploy this system as an online service.

1 Introduction

In this project, our focus is on creating a system that tackles the problem of multi-class image classification. For this purpose, we will mainly focus on food-101 Dataset which contains a huge set of labeled food images and use it for training our model. throughout this project, we're going to use the advantages of state-of-the-art convulotional architectures, and fine-tune it for this domain specific task.

NOTE: This project is different than the proposal we have made at the very beginning of the course, and that's because we found some difficulties due to its complexity in the given time constraints.

2 Related Work

The interest in image classification has increased noticeably in the era of deep learning, and the task of improving the efficiency of image classification has grown dramatically due to its wide range of applications in computer vision and industry in general. Recently, a wide variety of attempts to create state-of-the-art model has been conducted, and the emergence of convolutional neural networks has pushed these attempts significantly [1]. Accordindly, the success of success of AlexNet [3], paved the path for many works to improve its performance. Among them, four representative works are ZFNet [5], VGGNet [4], and ResNet [2]. In our work, we are going to take the advantages of these architectures, and apply them for a domain specific task (food classification), the architecture, the model, and the data set, are going to be described in the following sections.

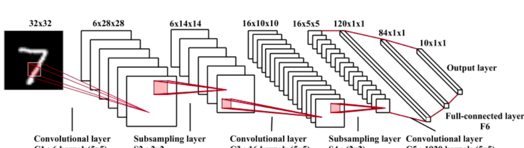

Figure 1. The main idea behind representation learning using CNNs

3 Data Set

For the purpose of multiclass food classification, we are going to use the food101, the data set contains a total 202000 images in total with 101 classes of food, the data is splitted into training and testing sets, the training set contains of 101000 images as well as 101000 images for the testing set.

4 Solution and System Architecture

Before going into the details of the neural network, let's provide a high level overview of the overall system. The system is going to use the image data set for training, and then use the obtained weights to perform the prediction task. The system uses a convolution down sampling architecture, that's going to learn the spatial features of input images, and outputs their vector representation, these vector embeddings are then going to be used to get the classification probabilities (using a softmax layer). A high-level overview of the system is presented by the diagram below,

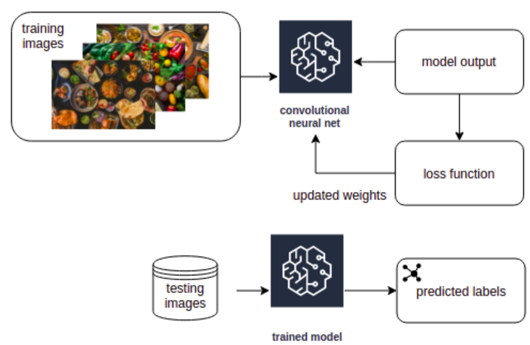

Figure 2. A high level overview of the system

The architecture of the neural network is illustrated in the following digram,

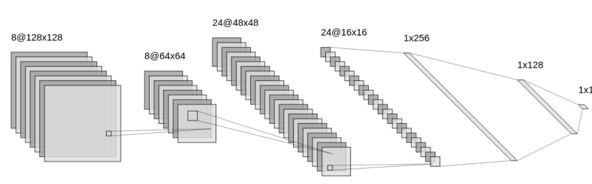

Figure 3. The architecture of the model

The system pipeline consists of the following components:

- ***Data pre-processing*** After the data is obtained and downloaded, we perform data pre-processing for the images, image pre-processing includes: normalization, cropping, and performing data augmentation before feeding the images to the data loader. These augmentation will encourage the model for better generalization.
- ***model training*** After pre-processing, we then feed the image to the model for training, the input data will go through the network layers, the output will be generated and the loss will be calculated to update the weights of the model.
- ***model testing*** After the training is done, the model can have representative vector embeddings of the images, and using these embeddings the model can obtain the probabilities to predict the corresponding class of the testing images.

4.1 Model deployment

We have deployed the model using streamlit, and we have conducted experiments using the GUI of the built web app. The following picture illustrates the interface of the web app,

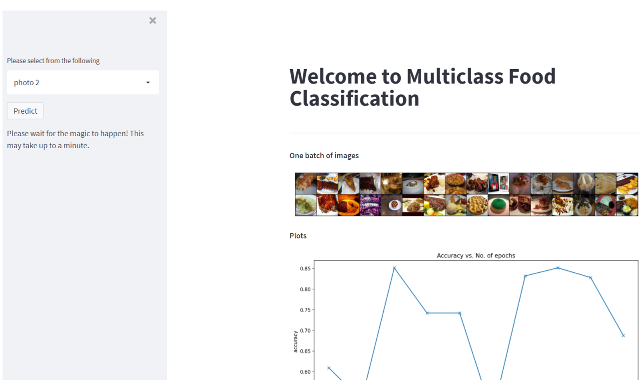

Figure 4. illustration of the deployed model

5 Results and evaluation

The model has been tested on the given test set, and fairly good results were obtained. The figure below shows the accuracy graph,

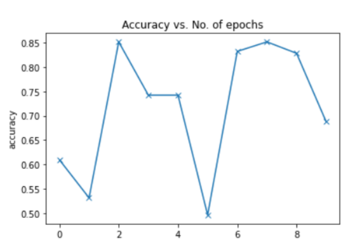

Figure 5. accuracy graph

And the following figure shows the loss function during training,

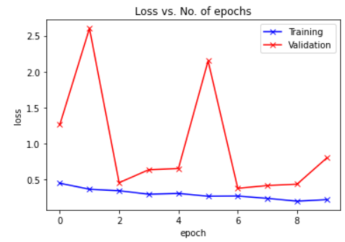

Figure 6. loss curves

6 Conclusion

CNNs are undoubtedly a breakthrough for image classification tasks. And using a model inspired by the state-of-the-art down sampling CNNs, we could create a system for food images classification with reasonably good accuracy. However, the results can be further enhanced by hyper-parameter tuning and refining the layers of the network.

References

[1] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, et al. 2018. Recent advances in convolutional neural networks. *Pattern Recognition* 77 (2018), 354–377.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* 25 (2012), 1097–1105.

[4] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[5] Matthew D Zeiler and Rob Fergus. 2014. Visualizing and understanding convolutional networks. In *European conference on computer vision*. Springer, 818–833.